Pingyue Zhang

Address: Shanghai Jiao Tong University, No.800 Dongchuan Road, Shanghai, 200240, China

Phone: +86 15821832601 | Email: williamzhangsjtu@sjtu.edu.cn | Website: https://williamzhangsjtu.github.io/

Research Interest

Multi-modal learning, Deep Learning, Self-supervised learning, Representation Learning

Background

M.Eng., Computer Science, *Shanghai Jiao Tong Univerisity, Shanghai, China* *Sept 2021 - Mar 2024*

- Grades (Cumulative GPA): 3.88/4.0
- Supervisor: Prof. Mengyue Wu

B.Eng. (Honor), Computer Science (IEEE Honor Class), *Shanghai Jiao Tong Univerisity, Shanghai, China* *Sept 2017 - June 2021*

- Grades (Cumulative Scores): 90/100, Ranking: 11/84

Honor & Awards

Scholarship

- HUAWEI Scholarship (RMB 5000, Sept 2019)
- Zhiyuan College Honors Scholarship (top 5%, RMB 5000 per year, 2017 - 2021)
- HYPERGRYPH Scholarship (RMB 5000, Sept 2022)

Travel Grant

- SIGMM Travel Grants for ACM MM 2021 ($1000)

Competition

- Silver model in IBM Call For Code Competition, Shanghai (RMB 10000, July 2019)

Research Experience

Research in Shanghai Jiao Tong University X-LANCE LAB *(Directed by Prof. Mengyue Wu)* *Sept 2019 - Present*

- **Automatic detection of disease like depression or Parkinson based on audio or text data.**
  - Develop a CBoW-like self-supervised method to pretrain a model for embedding extraction from audio, specifically designed for the purpose of depression detection.
  - Employ a pretrained model, originally designed for predicting emojis in social media text, to extract embeddings related to emojis from textual data for depression detection.
  - Develop a feature set derived from ASR transcripts of audio data to significantly enhance the classification performance for Parkinson's disease detection
  - Utilize extra emotion information to improve self-supervised contrastive method for depression detection.
- **Enhance the contrastive learning method to improve classification performance.**
  - Introduce an innovative contrastive loss that assigns varying weights to samples based on their label overlap with the anchor for better multi-label classification performance.
  - Propose a novel contrastive loss which leverages both label and semantic information from each sample to improve single-label classification performance.
- **Zero-shot learning.**
  - Utilize sound attributes generated by a large language model to train audio-text alignment, thereby improving zero-shot audio classification.

Publications

Zhang, P., *Wu, M., Dinkel, H., & Yu, K. (2021, October). Depa: Self-supervised audio embedding for depression detection. In Proceedings of the 29th ACM international conference on multimedia (pp. 135-143).*

- This paper proposes a CBoW-like self-supervised pretraining method. The pretrained model is used to extract segment-level embedding from audio to conduct depression detection.

Zhang, P., *Wu, M., Yu, K. (2023) ReCLR: Reference-Enhanced Contrastive Learning of Audio Representation for Depression Detection. Proc. INTERSPEECH 2023, 2998-3002, doi: 10.21437/Interspeech.2023-2474* (**Oral**)

- The paper utilizes an emotion recognition model to extract emotion information as reference features. A novel reference-enhanced contrastive learning is proposed to select fine-grained positive instances and allocate weight to negative instances, in order to enhance the depression detection performance.

Zhang, P., *Wu, M., & Yu, K. (2023). Using Emoji as an Emotion Modality in Text-Based Depression Detection. In National Conference on Man-Machine Speech Communication (pp. 59-67). Springer, Singapore.* (**Oral**)

- This paper focuses on utilizing emoji as an emotional modality to detect. In particular, the paper proposes to extract segment-level emotional information with model pretrained to predict emoji of text on social media.

ZHANG, P., *ZHANG, M., LIU, H., & Yang, Y. (2022). Prediction of Protein Subcellular Localization Based on Microscopic Images via Multi-Task Multi-Instance Learning. Chinese Journal of Electronics, 31(5), 888-896.*

- This paper proposes a multi-task learning strategy and mask generation to tackle multi-instance problem and enhance the prediction performance of protein subcellular localization.

Zhang, P., *Wu, M., & Yu, K. Automatic Parkinson's Speech Severity Prediction via Read Speech. In National Conference on Man-Machine Speech Communication.* (Accepted, **Oral, Best Paper Nominee**)

- This paper focuses on utilizing read speech to detect the speech severity of patients.

Xu, X., Zhang, Z., Zhou, Z., **ZHANG, P.,** *Xie, Z., Wu, M., & Zhu, K. Q. (2023). BLAT: Bootstrapping Language-Audio Pre-training based on AudioSet Tag-guided Synthetic Data. In Proceedings of the 31th ACM international conference on multimedia.*

- The paper utilizes audio captioning to generate text directly from audio and also proposes caption generation under the guidance of AudioSet tags.

Publications Under Review

A novel method that extends supervised contrastive learning to a multi-label setting. *(First Author, Submitted to AAAI-2024)*

- This paper proposes a novel supervised contrastive loss to deal with multi-label classification by adjusting weights based on how much label overlap one sample shares with the anchor.

Semantic-Enhanced Supervised Contrastive Learning*(First Author, Submitted to ICASSP-2024)*

- The paper proposes a novel supervised contrastive learning method which not only leverages label information but also utilizes semantic information to better select and weigh positives and negatives.

Zero-shot Classification Using Sound Attributes from Large Language Model*(Co-First Author, Submitted to ICASSP-2024)*

- This paper presents a novel approach for zero-shot audio classification using automatically generated sound attribute descriptions.

Skills

Language

- Native Mandarin, Fluent English (TOEFL iBT score: 105)

Programming language & Tools

- Python, C/C++; Pytorch, Git, Vim